Applied Soft Computing 72 (2018) 261–272

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Extended belief-rule-based system with new activation rule determination and weight calculation for classification problems

Long-Hao Yang [a,c], Jun Liu [d], Ying-Ming Wang [a,b,*], Luis Martínez [c]

[a] Decision Sciences Institute, Fuzhou University, Fuzhou, Fujian, PR China
[b] Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education, Fuzhou University, Fuzhou, Fujian, PR China
[c] Department of Computer Science, University of Jaén, Jaén, Spain
[d] School of Computing, Ulster University at Jorhanstown Campus, Northern Ireland, UK

ARTICLE INFO

Article history:
Received 3 October 2017
Received in revised form 12 July 2018
Accepted 2 August 2018
Available online 6 August 2018

Keywords:
Extended belief-rule-based system
Classification problem
Activation rule determination
Activation weight calculation
Multi-class

ABSTRACT

Among many rule-based systems employed to deal with classification problems, the extended belief-rule-based (EBRB) system is an effective and efficient tool and also has potentials in handing both quantitative and qualitative information under uncertainty. Despite many advantages, several drawbacks must be overcome for better applying the conventional EBRB system, including counterintuitive individual matching degrees, insensitivity to the calculation of individual matching degrees, and the inconsistency problem. Accordingly, by constructing the activation region of extended belief rules and revising the calculation formula of activation weights, the new procedures of activation rule determination and weight calculation are proposed to improve the conventional EBRB system, while the original procedures of rule inference and class estimation are retained from the conventional EBRB system. Nineteen classification datasets with different numbers of classes are studied to validate the efficiency and effectiveness of the proposed EBRB classification system compared with existing works. The comparison results demonstrate that the proposed EBRB classification system not only obtains a high accuracy better than the conventional EBRB system, but also has an excellent response time for classification. More importantly, the results derived from multi-class datasets show the significant performance of the proposed EBRB classification system compared with some state of art classification tools.

1. Introduction

Classification is a common and fundamental problem involved in various theoretical and practical applications, including, but not limited to, image processing [1], medical application [2], pattern recognition [3], and intrusion detection [4]. However, classification problems are not always easy to be solved due to the interconnected sophistications of the correlations attached to high dimensions [5] and the large number of samples [6]. Therefore, proposing an effective system or approach to handle classification problems is one of the most urgent challenges in these applications.

As one of the most visible and fastest growing branches of artificial intelligence (AI) [7], rule-based systems have been applied to handle classification problems in the past decades. The common rule-based systems include fuzzy rule-based (FRB) systems [8,9] and belief rule-based (BRB) systems [10–12]. Because of the capability for building a linguistic model interpretable to users and handling both quantitative and qualitative information derived from domain experts and mathematical models [13], these rule-based systems have been successfully utilized to deal with classification problems and have played an important role with some distinct advantages compared with some black-box classification tools, e.g. deep learning based classification tools [14,15].

However, many existing rule-based systems suffer from the combination explosion problem while dealing with high dimensional classification datasets, because of the rule generation that is required to cover all combinations of each alternative for each attribute [16]. For example, the establishment of fuzzy regions is the most important step to construct FRB systems but the number of fuzzy regions is always growing exponentially along with the increase of attributes and fuzzy numbers [8]. Meanwhile, the time complexity of generating a rule base usually imposes hard restrictions on the rule-based systems, mainly because the rule base needs to be trained iteratively and thus it is very expensive to train and re-train them [17]. For instance, parameter-learning is a

* Corresponding author at: Decision Sciences Institute, Fuzhou University, Fuzhou, Fujian, PR China. Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education, Fuzhou University, Fuzhou, Fujian, PR China.
E-mail address: ymwang@fzu.edu.cn (Y.-M. Wang).

https://doi.org/10.1016/j.asoc.2018.08.004
1568-4946/

well-known approach to train parameters for the BRB system. Actually, it is a time-consuming process to solve nonlinear programming problems [18,19], especially for the large number of training data.

Recently, the extended-rule-based (EBRB) system developed by Liu et al. [20] has three advantages compared with the FRB system and the BRB system.

(1) Belief structures are embedded into the consequent attribute and all antecedent attributes of an extended belief rule. Thus, both of fuzzy rules and belief rules are a special case of extended belief rules [21];
(2) The EBRB system is either a knowledge-driven or a data-driven, or combined decision model. It can be a data-driven model in the sense that extended belief rules can be generated from input-output data pairs [22];
(3) Now that the EBRB system can be a data-driven decision model, it is unnecessary to obtain the optimal parameters by using the parameter-learning which has been proven to be a time-consuming process [20].

Therefore, the EBRB system has the potential to be an effective and efficient tool to handle classification problems. However, several drawbacks were found in the procedures of activation rule determination and weight calculation in the conventional EBRB system. The most significant drawback is that the conventional EBRB system may suffer from the inconsistency problem in the conventional activation rule determination because two or more rules with different consequents are activated for the same input. Another two significant drawbacks were found in the conventional activation weight calculation, the conventional EBRB system may produce counterintuitive individual matching degrees while calculating activation weights and sometimes it is over insensitive to the calculation of individual matching degrees. These drawbacks must be overcome to more effectively utilize the conventional EBRB system in classification problems.

Motivated by these drawbacks, new procedures of activation rule determination and weight calculation are proposed.

For the new procedure of rule activation determination, the key idea is to construct activation regions for each extended belief rule and then use the activation regions to judge which extended belief rule should be activated. Since the activation regions can be constructed before utilizing the EBRB system to classify input data with a low requirement of real-time, utilizing activation regions can help the EBRB system determine consistent activation rules more efficiently.

For the new procedure of activation weight calculation, it aims to revise the original calculation formula of activation weights. Through mathematical derivations, it is easy to find the limitations of the conventional activation weight calculation. Actually, these limitations are the causes of counterintuitive results and insensitive calculations found in the conventional EBRB system. By considering the importance of referential values and the normalization of individual matching degrees, the new activation weight calculation can help the EBRB system calculate activation weights for activation rules more effectively.

In addition, the original procedures, including using the evidential reasoning (ER) algorithm [23,24] as the inference engine and the strategy of seeking maximum belief degree as class estimation, are retained from the conventional EBRB system. Finally, based on these procedures, a novel EBRB classification system is developed in this study.

In order to demonstrate the efficiency and effectiveness of the proposed EBRB classification system, a case study with nineteen classification datasets comprising two/three/multi-class are carried out to test the performance of the proposed EBRB classification system compared with existing works. Three aspects, namely, accuracy, failed data, and response time, are adopted to compare with other improved EBRB systems and popular machine learning approaches.

The remainder of the study is organized as follows: Section 2 briefly reviews the applicability and drawbacks of the conventional EBRB system in classification problems. Section 3 introduces and illustrates a novel EBRB classification system and the procedures of new rule activation determination and its weight calculation. Section 4 provides comparative case studies to demonstrate the efficiency and effectiveness of the proposed EBRB classification system, and the paper is concluded in Section 5.

2. Conventional EBRB system for classification problems

In this section, a well-known classification dataset is used to illustrate how the conventional EBRB system can handle classification problems and what significant drawbacks can be overcome to better develop a novel EBRB classification system.

2.1. Applicability of conventional EBRB system to classification problems

Suppose U_i is the ith ($i = 1,\ldots, M$) antecedent attribute in an EBRB, $A_{i,j}$ is the jth ($j = 1,\ldots, J_i$) referential value of the ith antecedent attribute, D is the consequent attribute in the EBRB, and D_n is the nth ($n = 1,\ldots, N$) referential value of the consequent attribute D. The kth ($k = 1,\ldots, L$) extended belief rule in the EBRB can thenbe written as:

$$R_k: IF\ U_1\ is\ \{(A_{1,j}, \alpha^k_{1,j}); j = 1, \ldots, J_1\} \wedge \cdots \wedge U_M\ is\ \{(A_{M,j}, \alpha^k_{M,j}); j = 1, \ldots, J_M\},\ THEN\ D\ is\ \{(D_n, \beta^k_n); n = 1, \ldots, N\} \quad (1)$$

where $\alpha^k_{i,j}$ ($0 \le \alpha^k_{i,j} \le 1$) and β^k_n ($0 \le \beta^k_n \le 1$) are the belief degrees of referential value $A_{i,j}$ and referential value D_n in the kth rule with $\sum_{j=1}^{J_i} \alpha^k_{i,j} \le 1$ and $\sum_{n=1}^{N} \beta^k_n \le 1$. In addition, the weight of the ith antecedent attribute and the kth rule are denoted as δ_i ($0 < \delta_i \le 1$) and θ_k ($0 < \theta_k \le 1$), respectively.

Let a classification dataset, Iris [25], illustrate the applicability of the conventional EBRB system to classification problems. For convenience, two antecedent attributes "*sepal length*" and "*sepal width*" are used to distinguish the species of iris. Three referential values of the two antecedent attributes and one consequent attribute are provided as follows:

$$U_1 = \{Low, Medium, High\} = \{A_{1,1}, A_{1,2}, A_{1,3}\} = \{4.3,\ 6.1,\ 7.9\} \quad (2a)$$

$$U_2 = \{Low, Medium, High\} = \{A_{2,1}, A_{2,2}, A_{2,3}\} = \{2.0,\ 3.2,\ 4.4\} \quad (2b)$$

$$D = \{Setosa, Versicolor, Virgica\} = \{D_1, D_2, D_3\} \quad (2c)$$

Suppose that one input-output data pair of dataset Iris, namely, {4.3, 3.2, *Virgica*}, can be transformed into a belief distribution format as introduced in [17].

$$\{\{(A_{1,1}, 1), (A_{1,2}, 0), (A_{1,3}, 0)\},\ \{(A_{2,1}, 0), (A_{2,2}, 1), (A_{2,3}, 0)\},\ Virgica\} \quad (3)$$

Since the input-output data pair is categorized as "*Virgica*", which means that the belief degree of being "*Virgica*" is 100% and

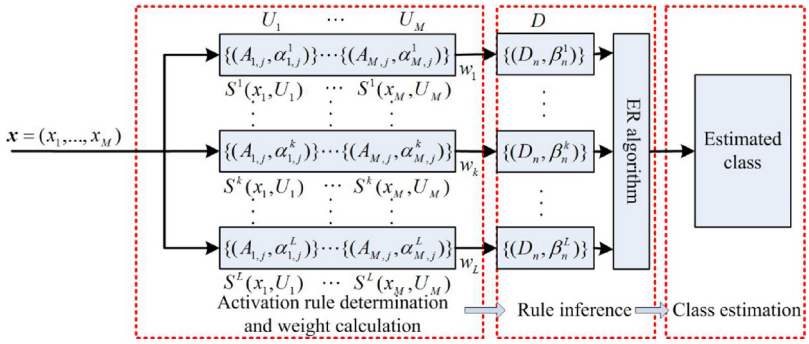

Fig. 1. Illustration of conventional EBRB system.

the other belief degrees are both 0%, the belief distribution of the output data can be written as follows:

$$\{(D_1, 0), (D_2, 0), (D_3, 1)\} \tag{4}$$

Consequently, an extended belief rule transformed from the input-output data pair can be represented as follows:

$$R_k: \textit{IF } U_1 \textit{ is } \{(A_{1,1}, 1), (A_{1,2}, 0), (A_{1,3}, 0)\} \wedge U_2 \textit{ is } \{(A_{2,1}, 0), (A_{2,2}, 1), (A_{2,3}, 0)\}, \textit{ THEN } D \textit{ is } \{(D_1, 0), (D_2, 0), (D_3, 1)\} \tag{5}$$

Following the above transformation, all input-output data pairs of dataset Iris can be transformed into extended belief rules, which further construct one EBRB system (see literature [20] for details). When an input data is provided for the EBRB system shown in Fig. 1, extended belief rules should be activated and can then be used to generate an estimated class for the input data.

From Fig. 1, the procedures of activation rule determination and weight calculation include many intermediate variables such as individual matching degrees $S^k(x_i, U_i)$ and activation weights w_k. Readers can refer to the literature [20] for complete details. To show the calculation process of these variables, the following three extended belief rules are used here:

$$R_1: \textit{IF } U_1 \textit{ is } \{(A_{1,1}, 0), (A_{1,2}, 0.2), (A_{1,3}, 0.8)\} \wedge U_2 \textit{ is } \{(A_{2,1}, 0.7), (A_{2,2}, 0.0), (A_{2,3}, 0.3)\}, \textit{ THEN } D \textit{ is } \{(D_1, 1), (D_2, 0), (D_3, 0)\} \tag{6a}$$

$$R_2: \textit{IF } U_1 \textit{ is } \{(A_{1,1}, 0), (A_{1,2}, 0.8), (A_{1,3}, 0.2)\} \wedge U_2 \textit{ is } \{(A_{2,1}, 0), (A_{2,2}, 0.7), (A_{2,3}, 0.3)\}, \textit{ THEN } D \textit{ is } \{(D_1, 0), (D_2, 1), (D_3, 0)\} \tag{6b}$$

$$R_3: \textit{IF } U_1 \textit{ is } \{(A_{1,1}, 0.5), (A_{1,2}, 0.5), (A_{1,3}, 0)\} \wedge U_2 \textit{ is } \{(A_{2,1}, 0.7), (A_{2,2}, 0.3), (A_{2,3}, 0)\}, \textit{ THEN } D \textit{ is } \{(D_1, 0), (D_2, 0), (D_3, 1)\} \tag{6c}$$

An input data $\boldsymbol{x}$ of the EBRB system is then supposed as follows:

$$\boldsymbol{x} = \{\{(Low, 1), (Medium, 0), (High, 0)\}, \{(Low, 1), (Medium, 0), (High, 0)\}\} \tag{7}$$

Consider that the weights of the three rules and two antecedent attributes are all set to 1. Hence, the individual matching degree of the two antecedent attributes and the activation weight of the three rules can be calculated by using Eqs. (10) and (11) shown in Section 3.1 and the calculated values are then listed in Table 1. Finally, the second and third rules are activated for the input data $\boldsymbol{x}$.

Table 1
Individual matching degrees and activation weights from conventional EBRB system.

Rule no. (R_k)	Individual matching degree ($S^k(x_i, U_i)$)		Activation weight (w_k)
	Sepal length (U_1)	Sepal width (U_2)	
1	−0.2961	0.5757	−0.1705
2	−0.2961	−0.2570	0.3110
3	0.2929	0.5757	0.6890

2.2. Related works of conventional EBRB system in classification problems

Since the EBRB system was shown to be applicable for handing classification problems, many attempts have been made to apply and improve the conventional EBRB system. For example, Calzada et al. [26] proposed the dynamic rule activation (DRA) method to dynamically adjust the set of activation rules, in which the essence of the DRA method is an iterative algorithm that needs to recalculate activation weight for each rule repeatedly. They showed that the DRA method can improve the accuracy and solve the inconsistency problem of the conventional EBRB system in the case of 22 classification datasets. Yang et al. [22] constructed the multi-attribute search framework (MaSF) based on the tree-based data structure, in which the MaSF contributes to decreasing the time complexity of the conventional EBRB system because there is no need to visit the entire EBRB to search for the activation rules. They concluded that the MaSF improves the accuracy and efficiency of the conventional EBRB system for 19 classification datasets. Yang et al. [21] introduced the data envelopment analysis (DEA) to reduce inefficient rules for the conventional EBRB system. They showed that 23%, 8%, 22%, 11%, and 9% rules can be reduced from the classification datasets Iris, Seeds, Ecoli, Diabetes, and Glass, respectively. Espinilla et al. [27] proposed the adaption of the EBRB system for human activity recognition in a smart environment. They suggested that the Hamming distance can improve the accuracy of the EBRB system for the binary classification problems. Recently, Yang et al. [28] proposed the consistency analysis-based rule activation (CABRA) method to define and select suitable activation rules, where the essence of the CABRA method is a linear optimization model. They showed that the CABRA method can solve the inconsistency problem and also make the conventional EBRB system have a better accuracy than that improved by the DRA method for 9 classification datasets.

Among the above attempts for improving the conventional EBRB system in classification problems, the DRA method and the CABRA method are the typical representatives because both of them aim at solving the inconsistency problem by embedding an activation mechanism into the process of determining activation rules, which is the same starting point to the present work. However, the essence

of the DRA method and the CABRA method decides that they need a lot of extra time to perform the iterative algorithm or the linear optimization model during determining activation rules for each input data. Clearly, this situation should be avoided as much as possible in the improved EBRB system, which is another starting point in the present work. As such, the case studies shown in Section 4 are mainly for comparative analysis based on the DRA method and CABRA method.

2.3. Drawbacks of conventional EBRB system in classification problems

As mentioned before, the conventional EBRB system suffers from some drawbacks while being used to handle classification problems. These drawbacks can be summarized as follows:

(1) The most significant drawback is that the EBRB system may suffer from the inconsistency problem, which will decrease the accuracy of the EBRB system because almost all rules must be activated no matter what input data is provided for the EBRB system. Clearly, Table 1 shows that except for the first rule which has an irrational activation weight, the other rules are all activated for the input data. Although many attempts, including the DRA method [26] and the CABRA method [28], were made to solve the inconsistency problem, they all failed to effectively construct an efficient procedure of activation rule determination for the EBRB system.
(2) Another significant drawback is that the EBRB system may produce counterintuitive individual matching degrees in the conventional procedure of activation weight calculation. As shown in Table 1, three individual matching degrees are negative values and these values may further produce negative activation weights for such as the first and second rules. Although the second rule has a positive activation weight, it is calculated from two negative individual matching degrees, namely −0.2961 and −0.2570. Such weights obviously make no sense and are irrational, which may restrict the application of the EBRB system.
(3) The last significant drawback is that the EBRB system is sometimes over insensitive to the calculation of individual matching degrees. As shown in Table 1, the individual matching degree of "*sepal width*" regarding the first rule is the same to that regarding the third rule. However, the belief distribution of "*sepal width*" regarding the first rule is $\{(A_{2,1}, 0.7), (A_{2,2}, 0.0), (A_{2,3}, 0.3)\}$ and that regarding the third rule is $\{(A_{2,1}, 0.7), (A_{2,2}, 0.3), (A_{2,3}, 0.0)\}$. In other words, two different rules produce a same individual matching degree for the input data. This drawback may also restrict the application of the EBRB system.

In the subsequent sections, it is necessary to develop the new procedures of activation rule determination and weight calculation for the EBRB system to overcome the above-mentioned drawbacks in classification problems.

3. New EBRB system for classification problems

This section introduces the conventional procedures of activation rule determination and weight calculation to investigate the causes of the drawbacks shown in Section 2.2. Based on these causes, new procedures of activation rule determination and weight calculation are proposed to improve the EBRB system. Besides, by using the conventional procedures of rule inference and class estimation, a novel EBRB classification system is developed to deal with classification problems.

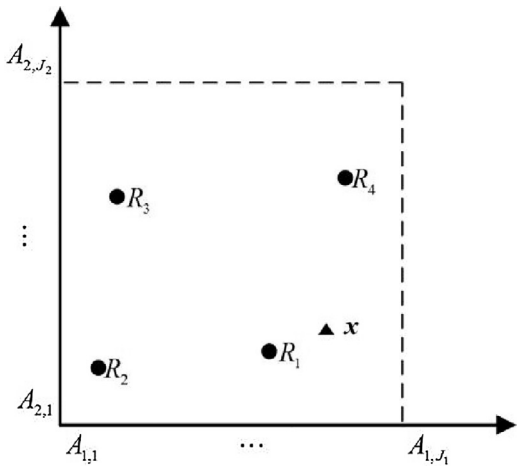

Fig. 2. Input region and conventional activation rule determination.

3.1. Analysis of conventional activation rule determination and weight calculation

In order to illustrate the conventional procedure of activation rule determination, consider that there is one classification example with two antecedent attributes, and each antecedent attribute has a set of referential values, namely $\{A_{1,1}, \ldots, A_{1,J_1}\}$ and $\{A_{2,1}, \ldots, A_{2,J_2}\}$, respectively. As such, the input space of the EBRB system can be represented by all lowest and highest referential values [29], as shown in Fig. 2.

From Fig. 2, there are four rules $\{R_1, R_2, R_3, R_4\}$ and one input data $\boldsymbol{x}$ in the input space of the EBRB system. To classify the input data $\boldsymbol{x}$, the rules located in the input space are directly used to calculate activation weights. In other words, all rules R_1, R_2, R_3, and R_4 should take part in the conventional procedure of activation weight calculation for the input data $\boldsymbol{x}$. Clearly, for the EBRB system, the cause of the inconsistency problem is due to lack of effective approach to screening consistent activation rules.

In order to illustrate the conventional procedure of activation weight calculation, consider that there is an input data $\boldsymbol{x}=(x_1, \ldots, x_M)$, where M is the number of antecedent attributes. Each input x_i is transformed into the following belief distribution [30]:

$$S(x_i) = \{(A_{i,j}, \alpha_{i,j}); j = 1, \ldots, J_i\} \tag{8}$$

where

$$\begin{aligned} &\alpha_{i,j} = \frac{u(A_{i,j+1}) - x_i}{u(A_{i,j+1}) - u(A_{i,j})} \text{ and } \alpha_{i,j+1} = 1 - \alpha_{i,j}, \text{ if } u(A_{i,j}) \le x_i \le u(A_{i,j+1}) \\ &\alpha_{i,k} = 0, \text{ for } k = 1, \ldots, J_i \text{ and } k \ne j, j+1. \end{aligned} \tag{9}$$

where $u(A_{i,j})$ is the utility value of the referential value $A_{i,j}$, $\alpha_{i,j}$ is the similarity degree to which the input x_i matches the referential value $A_{i,j}$, J_i is the number of referential values referred to the ith antecedent attribute.

The individual matching degree of the ith antecedent attribute in the kth rule is calculated by the following formula:

$$S^k(x_i, U_i) = 1 - \sqrt{\sum\nolimits_{j=1}^{Ji} (\alpha_{i,j} - \alpha_{i,j}^k)^2} \tag{10}$$

where $\alpha_{i,j}^k$ is the belief degree to which antecedent attribute U_i is evaluated to be the referential value $A_{i,j}$ in the kth rule, $S^k(x_i,U_i)$ is the individual matching degree of the input x_i to the antecedent attribute U_i in the kth rule.

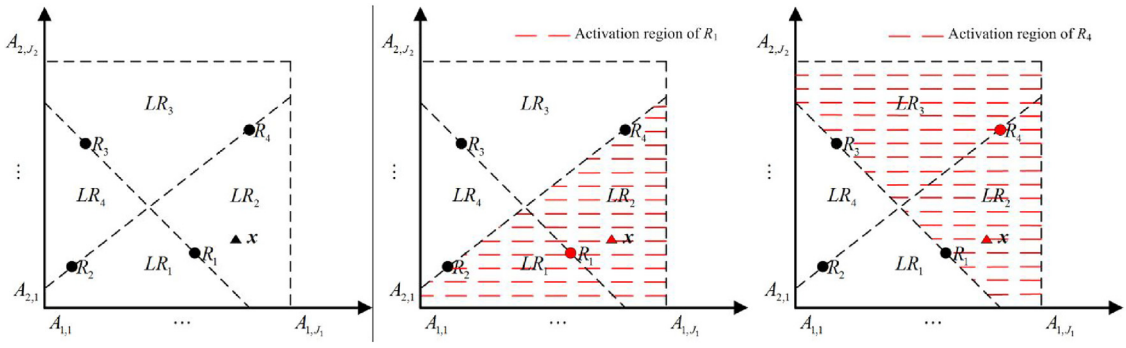

Fig. 3. Activation regions and new rule activation determination. (a) Activation regions (b) Activation region of R_1 (c) Activation region of R_4.

The activation weight for the kth rule is then calculated by the following formulas:

$$w_k = \frac{\theta_k \prod_{i=1}^{M} \left(S^k(x_i, U_i)\right)^{\bar{\delta}_i}}{\sum_{l=1}^{L} \left(\theta_l \prod_{j=1}^{M} \left(S^l(x_j, U_j)\right)^{\bar{\delta}_j}\right)}, \quad \bar{\delta}_i = \frac{\delta_i}{\max_{j=1,\ldots,M}\{\delta_j\}} \tag{11}$$

where θ_k is the weight of the kth rule; δ_i is the weight of the ith antecedent attribute; and w_k is the activation weight of the kth rule.

To investigate the causes of the drawbacks found in the conventional EBRB system, the formula derivations shown in Eqs. (12) and (13) are used to analyze the conventional procedure of activation weight calculation.

Firstly, according to Eq. (10), the calculation formula of individual matching degrees can be deduced as follows:

$$\begin{aligned} S^k(x_i, U_i) &= 1 - \sqrt{\textstyle\sum_{j=1}^{Ji} \left(\alpha_{i,j} - \alpha_{i,j}^k\right)^2} \geq 1 - \sqrt{\textstyle\sum_{j=1}^{Ji} |\alpha_{i,j} - \alpha_{i,j}^k|} \\ &\geq 1 - \sqrt{\textstyle\sum_{j=1}^{Ji} (|\alpha_{i,j}| + |\alpha_{i,j}^k|)} \\ &= 1 - \sqrt{\textstyle\sum_{j=1}^{Ji} \alpha_{i,j} + \sum_{j=1}^{Ji} \alpha_{i,j}^k} \geq 1 - \sqrt{2} \end{aligned} \tag{12}$$

It is obvious from Eq. (12) that a necessary normalization is neglected in Eq. (10), so that the lower bound of individual matching degrees is 1-$\sqrt{2}$. This is why the conventional EBRB system produces counterintuitive individual matching degrees.

Secondly, when Eq. (10) is regarded as a special weighted calculation formula, it can be written as follows:

$$S^k(x_i, U_i) = 1 - \sqrt{\sum_{j=1}^{Ji} \left(\alpha_{i,j} - \alpha_{i,j}^k\right)^2} = 1 - \sqrt{\sum_{j=1}^{Ji} \varepsilon_{i,j} \left(\alpha_{i,j} - \alpha_{i,j}^k\right)^2} \tag{13}$$

where $\varepsilon_{i,j}$ is the weight of the referential value $A_{i,j}$ and its value is equal to 1, namely, $\varepsilon_{i,j} = 1$ ($i = 1,\ldots, M$; $j = 1,\ldots, J_i$), for the conventional procedure of activation weight calculation.

It is clear from Eq. (13) that the difference between referential values is neglected in Eq. (10), so that the conventional EBRB system is sometimes over insensitive to the calculation of individual matching degrees.

3.2. New activation rule determination based on the activation region of extended belief rules

The new procedure of rule activation determination divides the input space of the EBRB system into four two-dimensional local regions using the lines which are made by two rules, as shown in Fig. 3(a), where LR_i represents the ith ($i = 1,\ldots, 4$) local region. The following definition is given to define the activation region of extended belief rules:

Definition 1. Suppose the input space of an EBRB system is decomposed into multiple local regions according to the lines which are made by extended belief rules. Thus, the activation region of the kth ($k = 1,\ldots, L$) extended belief rule is defined as the region that: 1) is formed by any adjacent local regions; 2) only includes the kth extended belief rule.

From Definition 1, if an input data falls into an activation region, the extended belief rule located in the activation region is necessary to be activated for the input data. For example on Fig. 3(a)–(c), Fig. 3(a) shows that the activation region of the rules R_1, R_2, R_3, and R_4 is $\{LR_1, LR_2\}$, $\{LR_1, LR_4\}$, $\{LR_3, LR_4\}$, and $\{LR_2, LR_3\}$, respectively. As shown in Fig. 3(b) and (c), if the input data $\boldsymbol{x}$ falls into the local region LR_2, then only rules R_1 and R_4 are activated and integrated to classify the input data $\boldsymbol{x}$.

To construct activation regions, below steps in the case of the kth ($k = 1,\ldots, L$) extended belief rule should be done:

Step 1. Express extended belief rules in the form of tuple [26]. Suppose $u(A_{i,j})$ is the utility value of the referential value $A_{i,j}$ ($i = 1,\ldots, M$; $j = 1,\ldots, J_i$) in the ith antecedent attribute, M is the number of antecedent attributes, J_i is the number of referential values in the ith antecedent attribute. Thus, the kth rule can be expressed by the following formulas:

$$T(R_k) = < x_1^k, \ldots, x_M^k > \tag{14}$$

where

$$x_i^k = \sum_{j=1}^{J_i} \alpha_{i,j}^k u(A_{i,j});\ i = 1, \ldots, M \tag{15}$$

where $\alpha_{i,j}^k$ is the belief degree of the referential value $A_{i,j}$ in the kth rule.

Step 2. Calculate the absolute distance of each antecedent attribute between the kth rule and the other rules $\{R_j; j = 1,\ldots, L; j \neq k\}$ in the form of tuple. For the sake of clarity, these distances are expressed by the following matrix:

$$D(R_k) = \begin{bmatrix} |x_1^k - x_1^1| & \cdots & |x_1^k - x_1^j| & \cdots & |x_1^k - x_1^L| \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ |x_M^k - x_M^1| & \cdots & |x_M^k - x_M^j| & \cdots & |x_M^k - x_M^L| \end{bmatrix} \tag{16}$$

Step 3. Construct lines that separate the kth rule from the others $\{R_j; j = 1,\ldots, L; j \neq k\}$ in the input space of the conventional EBRB system. Suppose that the line through $R_l \in \{R_j; j = 1,\ldots, L; j \neq k\}$ is able to separate the kth rule and the other rules $\{R_j; j = 1,\ldots, L; j \neq k\}$. In other words, there exists a set of weights $\{\omega_1^l, \ldots, \omega_M^l\}$ for the rule R_l to ensure the minimal weighted distance $\sum_{i=1}^{M} \omega_i^l |x_i^k - x_i^l|$ comparing to other rules $\{R_j; j = 1,\ldots, L; j \neq l; j \neq k\}$. Without loss of generality, it is assumed that the minimal weighted distance is 1.

The set of weights for the rule R_l can be obtained by the following linear programming model:

$$\begin{aligned} &\max\ \mu_l \\ &s.t.\ \textstyle\sum_{i=1}^{M}\omega_i^l|x_i^k - x_i^j| \geq \mu_l;\ j = 1, \ldots, L; j \neq l; j \neq k \\ &\quad \textstyle\sum_{i=1}^{M}\omega_i^l|x_i^k - x_i^l| = 1 \\ &\quad \omega_i^l \geq 0; i = 1, \ldots, M \end{aligned} \tag{17}$$

where μ_l is the maximum weighted distance obtained from the minimal weighted distance of the rules $\{R_j; j = 1,\ldots, L; j \neq l; j \neq k\}$. If $\mu_l < 1$, it means that the line through R_l fails to separate the kth rule and the other rules $\{R_j; j = 1,\ldots, L; j \neq k\}$; Otherwise, the set of weights $\{\omega_1^l, \ldots, \omega_M^l\}$ can be used to construct a line for the kth rule.

Step 4. Construct the activation region of the kth rule using the set of weights $\{\omega_1^l, \ldots, \omega_M^l\}$. The activation region of the kth rule can be expressed by the following formula:

$$Q(R_k) = \{(x_1, \ldots, x_M) \left| \sum_{i=1}^{M}\omega_i^l|x_i^k - x_i| \leq 1, l = 1, \ldots, L^k\right\} \tag{18}$$

where x_i is the input variable in the ith antecedent attribute; L^k is the number of rules that can be used to construct the line for the kth rule; $Q(R_k)$ is the activation region of the kth rule.

For the abovementioned steps, the following remarks are provided:

Remark 1. The activation region of each extended belief rule can be constructed by an offline process because it is completely independent of the procedures of activation rule determination and weight calculation. Hence, comparing with the existing activation rule determination, such as the DRA method and the CABRA method, the new rule activation determination is high-efficient due to the fact that the construction of activation regions has no effect on the time complexity of EBRB system.

Based on the activation region $Q(R_k)$, the new procedure of rule activation determination can be described as the following formula:

$$RA(Q(R_k), \mathbf{x}) = sign(\min_{1\leq l\leq L^k}(\sum_{i=1}^{M}\omega_i^l|x_i^k - x_i| - 1)) \tag{19}$$

where $\boldsymbol{x}$=(x_1,. . ., x_M) is the input data of EBRB system, $RA(Q(R_k), \boldsymbol{x})$ is the evaluation function of activating the kth rule. If $RA(Q(R_k), \boldsymbol{x})$ = -1, it means that the activation weight of the kth rule should be calculated for the input data $\boldsymbol{x}$; Otherwise, the rule should not be activated.

Remark 2. The new procedure of rule activation determination must be able to overcome the inconsistency problem found in the conventional EBRB system, because the activation region of each extended belief rule can be used to effectively determine consistent activation rules for each input data.

Remark 3. When input data are provided for an EBRB system, it is linear time complexity to determine activation rules. The time complexity of the rule activation determination procedure is also linearly increasing as the number of input data rises. Therefore, the new procedure of rule activation determination is an efficient procedure for the EBRB system.

3.3. New activation weight calculation by considering the weight of referential values and normalization

The new procedure of activation weight calculation assumes that there is an input data vector $\boldsymbol{x}$=(x_1,. . ., x_M) and a set of utility values $\{u(A_{i,j}); i = 1,\ldots, M; j = 1,\ldots, J_i\}$ regarding the referential value $A_{i,j}$, where the relationship between adjacent utility values in the ith antecedent attribute is $u(A_{i,j})<u(A_{i,j+1})$ (j = 1,. . .,J_i-1), M is the number of antecedent attributes, and J_i is the number of referential values referred to the ith antecedent attribute.

Hence, based on those utility values, the weight of the referential value $A_{i,j}$ is calculated as

$$\varepsilon_{i,j} = e^s,\ s = \frac{u(A_{i,j}) - u(A_{i,1})}{u(A_{i,J_i}) - u(A_{i,1})} \tag{20}$$

By considering the weight $\varepsilon_{i,j}$ and the normalization process, new individual matching degree of the ith antecedent attribute in the kth rule is calculated by

$$\bar{S}^k(x_i, U_i) = 1 - \sqrt{\frac{\sum_{j=1}^{Ji}\varepsilon_{i,j}(\alpha_{i,j} - \alpha_{i,j}^k)^2}{\varepsilon_{i,J_i-1} + \varepsilon_{i,J_i}}}, \tag{21}$$

where $\alpha_{i,j}^k$ is the belief degree to which the antecedent attribute U_i is evaluated to be the referential value $A_{i,j}$ in the kth rule, $\alpha_{i,j}$ is the similarity degree to which the input x_i matches the referential value $A_{i,j}$.

Based on the new individual matching degree, new activation weight of the kth rule is calculated by using Eq. (11) with the new individual matching degree, namely:

$$\bar{w}_k = \frac{\theta_k\prod_{i=1}^{M}\left(\bar{S}^k(x_i, U_i)\right)^{\bar{\delta}_i}}{\sum_{l=1}^{L}\left(\theta_l\prod_{j=1}^{M}\left(\bar{S}^l(x_j, U_j)\right)^{\bar{\delta}_j}\right)},\ \bar{\delta}_i = \frac{\delta_i}{\max_{j=1,\ldots,M}\{\delta_j\}} \tag{22}$$

where θ_k is the weight of the kth rule; δ_i is the weight of the ith antecedent attribute.

Remark 4. The new procedure of activation weight calculation overcomes the drawbacks of conventional EBRB system because of the following three reasons:

(1) As shown in Eq. (20), the weight of referential values can avoid $\varepsilon_{i,j} = 0$ (i = 1,. . ., M; j = 1,. . ., J_i) because the function e^s is a computation component to the weight $\varepsilon_{i,j}$ and the output of this function is bigger than 0 for any input data.
(2) As shown in Eq. (21), the weight of referential values is used to calculate individual matching degrees and these weights can illustrate different importance for different referential values.
(3) As shown in Eq. (21), the value range of individual matching degrees is [0,1] since two maximal weights of referential values are regarded as a denominator to normalize individual matching degrees.

Table 2
Individual matching degrees and activation weights from new activation weight calculation (the corrected values are highlighted in boldface).

Rule no. (R_k)	Individual matching degree ($\bar{S}^k(x_i, U_i)$)		Activation weight ($\bar{w}_k$)
	Sepal length (U_1)	Sepal width (U_2)	
1	**0.1985**	**0.7232**	**0.2037**
2	**0.2961**	**0.3144**	0.1321
3	0.6106	**0.7664**	0.6642

Table 2 shows the results obtained from the new activation weight calculation while $\{u(A_{1,1}), u(A_{1,2}), u(A_{1,3})\}$ to "*Sepal length*" is {4.3, 6.1, 7.9} and $\{u(A_{2,1}), u(A_{2,2}), u(A_{2,3})\}$ to "*Sepal width*" is {2.0, 3.2, 4.4}.

By comparing with Table 1, it is clear from the values marked as bold in Table 2 that the new procedure of activation weight calculation can overcome the drawback of counterintuitive individual matching degrees, e.g., new individual matching degrees and activation weights are all positive values, and the drawback

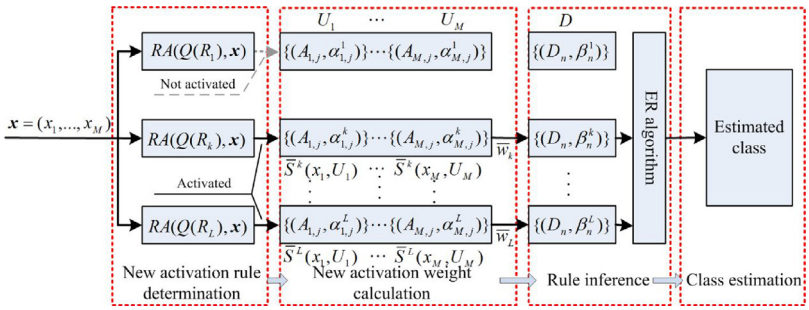

Fig. 4. Illustration of the proposed EBRB classification system.

of insensitive to the calculation of individual matching degrees, e.g., $\bar{S}^1(x_2,U_2) \neq \bar{S}^3(x_2,U_2)$ in Table 2 is more reasonable than $S^1(x_2,U_2)=S^3(x_2,U_2)$ in Table 1.

3.4. EBRB classification system with new activation rule determination and weight calculation

Based on the new procedures of activation rule determination and weight calculation, a novel EBRB classification system is developed in this section, as shown in Fig. 4.

There are two main processes in the proposed EBRB classification system. The first process is the new procedures of activation rule determination and weight calculation, these procedures aim to judge which rule should be activated and how to calculate activation weights. The second process is the original procedures of conventional EBRB system and their roles are to integrate activation rules using the ER algorithm and estimate the final class based on inference results. Correspondingly, four steps are given as follows:

Step 1. Activation rule determination based on the activation region of each extended belief rule shown in Section 3.2. In order to ensure the efficiency of the proposed EBRB classification system, the activation region of all extended belief rules is constructed in advance. After that, for the given input data $\boldsymbol{x}$, the proposed EBRB classification system can efficiently determine activation rules based on the evaluation function shown in Eq. (19).

Step 2. Activation weight calculation for each activation rule showed in Section 3.3. In order to calculate activation weights for each activation rules, the first step is to transform the input data $\boldsymbol{x}$ into the belief distribution by using Eqs. (8) and (9); the second step is to calculate individual matching degrees for each antecedent attribute by using Eqs. (20) and (21); and finally the third step is to calculate activation weights for each activation rule by using Eq. (22).

Step 3. Rule inference using the ER algorithm. After calculating activation weights, all activation rules should be integrated using the following analytical ER algorithm [23,24]:

$$\beta_n=\frac{\prod_{k=1}^{L^*}\left(\bar{w}_k\beta_n^k+1-\bar{w}_k\sum_{i=1}^{N}\beta_i^k\right)-\prod_{k=1}^{L^*}\left(1-\bar{w}_k\sum_{i=1}^{N}\beta_i^k\right)}{\sum_{i=1}^{N}\prod_{k=1}^{L^*}\left(\bar{w}_k\beta_i^k+1-\bar{w}_k\sum_{j=1}^{N}\beta_j^k\right)-(N-1)\prod_{k=1}^{L^*}\left(1-\bar{w}_k\sum_{j=1}^{N}\beta_j^k\right)-\prod_{k=1}^{L^*}(1-\bar{w}_k)} \tag{23}$$

where $\bar{w}_k$ is the new activation weight of the kth activation rule; β_n^k is the belief degree of the class D_n in the kth activation rule; L^* is the number of activation rules; and β_n is the integrated belief degree of the class D_n to classify the input data $\boldsymbol{x}$.

Step 4. Class estimation based on the inference result. The class estimation is to seek the biggest belief degree from the inference results to determine the final estimated class by the following formulas:

$$f(\boldsymbol{x})=D_t,\ t=\arg\max_{n=1,\ldots,N}(\beta_n) \tag{24}$$

Table 3
Statistics on nineteen classification datasets.

No.	Dataset	No. of samples	No. of attributes	No. of classes
1	Mammographic	830	5	2
2	Banana	5300	2	2
3	Titanic	2201	3	2
4	Banknote	1372	4	2
5	Phoneme	5404	5	2
6	Diabetes	393	8	2
7	Cancer	569	30	2
8	Iris	150	4	3
9	Seeds	210	7	3
10	Wine	178	13	3
11	Knowledge	403	5	4
12	Car evaluation	1728	6	4
13	Pageblocks	5473	10	4
14	Nursery	1296	8	5
15	Shuttle	2175	9	5
16	Glass	214	9	6
17	Red wine	1599	11	6
18	Ecoli	336	7	8
19	Yeast	1484	8	10

4. Case study

Nineteen classification datasets obtained from the University of California at Irvine [25] are studied to validate the efficiency and effectiveness of the proposed EBRB classification system. Table 3 summarizes the number of attributes, classes, and samples. For each classification dataset, the experiment results are obtained under the assumption that each antecedent attribute has five referential values and their utility values are generated uniformly from the upper and lower boundaries of each antecedent attribute. Furthermore, the 10/2-fold cross-validation [31] is performed and the proposed EBRB classification system is implemented in Microsoft Visual C++ on Intel (R) Core (TM) i5-4300U CPU @ 1.90 GHz with Windows 7.

4.1. Comparative analysis with the conventional EBRB systems

Based on the number of classes, the nine small scale classification datasets with the sample range [150, 830] are divided into three groups, namely the group of two-class datasets shown in Table 4, three-class datasets shown in Table 5, and multi-class datasets shown in Table 6. For five independent runs with 10-fold cross-validation, the sample of each dataset is randomly partitioned into 10 equal sized subsamples with 9 subsamples as a training dataset and the remaining subsample as a testing dataset. In addition, the conventional EBRB system [20] and its improvements, including the improved EBRB system by the DRA method [26] and the CABRA method [28], are used to compare the proposed EBRB classification system, where all these EBRB systems are abbreviated as C-EBRB, DRA-EBRB and CABRA-EBRB in Tables 4–6, respectively.

The average results obtained from 5 independent runs for each classification dataset are measured with: 1) Accuracy, which shows the percentage of testing data correctly classified by the EBRB system; 2) Failed data, which shows the number of testing data where the EBRB system cannot produce any estimated class due to lack of activation rules; and 3) Response time, which shows the running time of the EBRB system to classify each input data in millisecond.

For the group of two-class datasets shown in Table 4, CABRA-EBRB has the best accuracies in two of three datasets, which are 97.01% and 76.34% obtained from datasets Cancer and Diabetes, respectively. Despite this fact, it is still possible to see some considerable improvements in accuracies. For example, the accuracies of the proposed EBRB classification system are all better than those of the C-EBRB and DRA-EBRB. In terms of failed data, except for the results of C-EBRB obtained from datasets Cancer and Mammographic, the results of the other EBRB systems are 0 for different datasets. It is proved that the proposed EBRB classification system is the same as the other improved EBRB systems and can effectively activate rules to classify input data. In terms of response time, the results show that the CABRA-EBRB is the most time-consuming system compared with the others, and the proposed EBRB classification system needs more response time than the C-EBRB and DRA-EBRB but there is no much difference.

For the group of three-class datasets, Table 5 shows that the accuracies of the proposed EBRB classification system start being

Table 4
Comparisons of average results for two-class datasets (the best results are highlighted in boldface).

Dataset (No. of samples/classes)	Criterion	C-EBRB [26]	DRA-EBRB [26]	CABRA-EBRB [28]	This study
Cancer (569/2)	Accuracy (%)	94.59	94.61	**97.01**	95.47
	Failed data	8.8	0	0	0
	Response time (ms)	**27.5**	29.6	>68.5	35.2
Diabetes (393/2)	Accuracy (%)	73.39	71.44	**76.34**	75.98
	Failed data	0	0	0	0
	Response time (ms)	**6.0**	7.0	>28.8	7.1
Mammographic (830/2)	Accuracy (%)	77.64	78.39	79.52	**79.57**
	Failed data	0.1	0	0	0
	Response time (ms)	7.3	**6.8**	>11.0	7.5
Average	Accuracy (%)	81.87	81.48	**84.29**	83.67
	Failed data	3.0	0	0	0
	Response time (ms)	**13.6**	14.4	>36.1	16.6

Table 5
Comparisons of average results for three-class datasets (the best results are highlighted in boldface).

Dataset (No. of samples/classes)	Criterion	C-EBRB [26]	DRA-EBRB [26]	CABRA-EBRB [28]	This study
Seeds (210/3)	Accuracy (%)	87.04	92.02	92.38	**93.24**
	Failed data	0	0	0	0
	Response time (ms)	**2.8**	**2.8**	>14.8	3.4
Iris (150/3)	Accuracy (%)	95.20	95.50	**96.00**	95.73
	Failed data	0	0	0	0
	Response time (ms)	>1.6	>1.6	>15.7	**>1.2**
Wine (178/3)	Accuracy (%)	96.32	96.46	96.63	**97.87**
	Failed data	0	0	0	0
	Response time (ms)	**4.1**	8.1	>26.5	4.8
Average	Accuracy (%)	92.85	94.66	95.00	**95.61**
	Failed data	0	0	0	0
	Response time (ms)	**>2.8**	>4.2	>19.0	>3.1

Table 6
Comparisons of average results for multi-class datasets (the best results are highlighted in boldface).

Dataset (No. of samples/classes)	Criterion	C-EBRB [26]	DRA-EBRB [26]	CABRA-EBRB [28]	This study
Knowledge (403/4)	Accuracy (%)	75.07	80.71	89.33	**92.01**
	Failed data	0	0	0	0
	Response time (ms)	4.5	6.8	>21.7	**3.8**
Glass (214/6)	Accuracy (%)	51.43	69.65	72.90	**75.51**
	Failed data	3.55	0	0	0
	Response time (ms)	4.6	9.2	>28.5	**4.2**
Ecoli (336/8)	Accuracy (%)	33.72	83.76	85.42	**87.14**
	Failed data	1.00	0	0	0
	Response time (ms)	**4.3**	6.5	>27.8	4.9
Average	Accuracy (%)	53.41	78.04	82.55	**84.89**
	Failed data	1.52	0	0	0
	Response time (ms)	4.5	7.5	>26.0	**4.3**

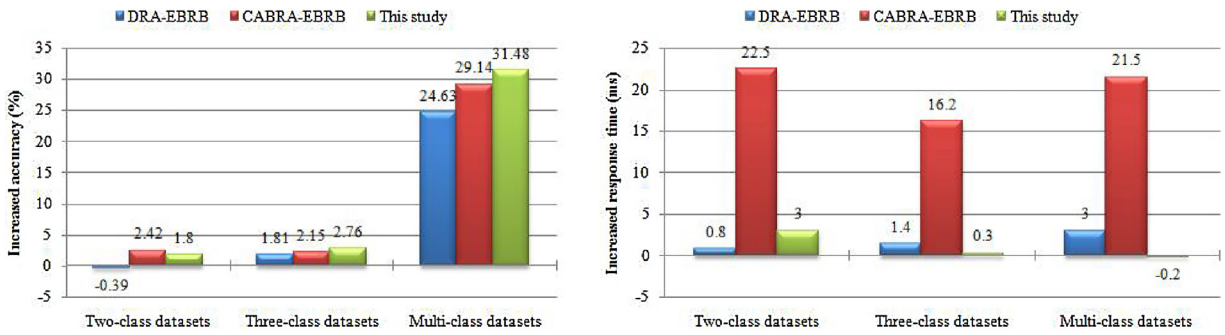

Fig. 5. Comparisons of increased accuracy and response time for different improved EBRB systems.
(a) Increased accuracy in different datasets (b) Increased response time in different datasets.

better comparing to others. In such group of datasets, the new procedure of rule activation can contribute to that more consistent activation rules are used to classify input data and it thus becomes a more important procedure than in the group of two-class datasets. Although for dataset Iris the accuracy improves just slightly and worse than the CABRA-EBRB, in the remaining datasets Seeds and Wine the accuracies improve very significantly and are better than the others. In addition, the proposed EBRB classification system takes the minimum response time to classify each input data for dataset Iris, where the minimum running time means that the proposed EBRB classification system is possible to improve the computing efficiency of the procedure of rule inference owning to fewer activation rules. Meanwhile, the considerable good efficiency of the proposed EBRB classification system is further demonstrated by comparing the response time of the DRA-EBRB and CARBA-EBRB. Taking dataset Wine for example, the response time derived from the proposed EBRB classification system is less than that from the DRA-EBRB and further less than that from the CABRA-EBRB.

For the group of multi-class datasets, it is clear from Table 6 that the proposed EBRB classification system can produce the best accuracies for all classification datasets and these accuracies are significantly better than the accuracies of the C-EBRB. Taking dataset Ecoli for example, the accuracy of the proposed EBRB classification system has been improved 258.4% compared with that of C-EBRB. Beyond that, the proposed EBRB classification system can classify each input data using the minimal response time, especially for datasets Glass and Knowledge where the response time of the proposed EBRB classification system are less than those of C-EBRB. In terms of failed data, except for 3.55 and 1 in datasets Glass and Ecoli for the C-EBRB, the other improved EBRB systems such as DRA-EBRB, CABRA-EBRB, and the proposed EBRB classification can effectively avoid undesired situation where none of rules are activated for special input data.

In order to further compare the proposed EBRB classification system with other improved EBRB systems including the DRA-EBRB and CABRA-EBRB, Fig. 5 shows their increased accuracies and response time in comparison with the conventional EBRB system based on the average results of two/three/multi-class datasets.

According to Fig. 5, some preliminary conclusions can be summarized as follows:

(1) Compared with the DRA-EBRB, the accuracies of the proposed EBRB classification system are always better than those of the conventional EBRB system.
(2) Compared with the CABRA-EBRB, the response time of the proposed EBRB classification system are close to the conventional EBRB system.
(3) For the multi-class datasets, the proposed EBRB classification system is the only system that can not only increase the accuracy of the conventional EBRB system, but also decreases the response time of the conventional EBRB system.

4.2. Comparative analysis under larger scale classification datasets

Consider that the comparative analyses in Section 4.1 are based on the small scale classification datasets with the sample range [150, 830]. Another ten larger scale classification datasets, which belong to the sample range [1296, 5473], are applied to compare the proposed EBRB classification system with other EBRB systems, including C-EBRB and DRA-EBRB, where the CABRA-EBRB does not involve in the comparative analysis based on these ten datasets because Tables 4–6 and Fig. 5 show that the CABRA-EBRB is much more time-consuming. Additionally, the results of Table 7 are obtained from 2-fold cross-validation for each dataset and are measured with: 1) Accuracy; 2) Distance, which shows the minimum attribute distance between activation rules and input data in terms of the single attribute value [22]; 3) Response time; 4) Activation ratio, which shows the average percentage of rules that is activated for each input data.

As we can observe in Table 7, some preliminary conclusions are summarized as follows:

(1) The proposed EBRB classification system has a better accuracy in two of four two-class datasets and all six multi-class datasets, e.g. it is clear from Table 7 that only the 98.83% and 78.06% accuracies of two-class datasets Banknote and Titanic obtained from the proposed EBRB classification system are worse than the 99.93% and 78.83% accuracies obtained from DRA-EBRB. For the remaining eight classification datasets, the accuracy of the proposed EBRB classification system outperforms all listed studies. Consequently, the average accuracy of the proposed system is better than that of C-EBRB and DRA-EBRB for the ten classification datasets.
(2) Apart from the dataset Yeast, the proposed EBRB classification system has a smaller distance between activation rules and input data, which means that the proposed EBRB classification system tends to activate the rule closer to input data, leading to more consistent activation rules and better accuracy for the proposed EBRB classification system. Clearly, the significant improvement is more likely to be found in the multi-class datasets because it is easier to activate inconsistent rules for an EBRB system while the dataset has more classes.
(3) The proposed EBRB classification system is completely better than DRA-EBRB but partially better than C-EBRB in terms of response time. For example, the average response time of the proposed EBRB classification system (5.6 ms) is more than that of C-EBRB (2.83 ms) but much less than that of DRA-EBRB (22.4 ms). The reason for this situation is that, as we can

Table 7
Comparisons of average results for larger scale datasets (the best results are highlighted in boldface).

Dataset (No. of samples/classes)	Criterion	C-EBRB	DRA-EBRB	This study
Banknote (1372/2)	Accuracy (%)	91.98	**99.93**	98.83
	Distance	106.467	55.858	**4.597**
	Response time (ms)	1.3	7.1	**1.1**
	Activation ratio (%)	100	51.62	**2.02**
Banana (5300/2)	Accuracy (%)	63.43	87.06	**87.28**
	Distance	630.752	609.177	**6.498**
	Response time (ms)	2.7	13.9	**2.4**
	Activation ratio (%)	100	97.64	**0.17**
Titanic (2201/2)	Accuracy (%)	76.74	**78.83**	78.06
	Distance	31.174	0.362	**0**
	Response time (ms)	**1.4**	8.7	2.0
	Activation ratio (%)	100	80.86	**48.26**
Phoneme (5404/2)	Accuracy (%)	71.89	76.02	**85.23**
	Distance	216.286	149.747	**10.109**
	Response time (ms)	**4.6**	34.5	8.4
	Activation ratio (%)	100	62.65	**1.42**
Car evaluation (1728/4)	Accuracy (%)	78.13	82.58	**90.22**
	Distance	75.382	18.563	**0**
	Response time (ms)	1.8	12.1	**1.3**
	Activation ratio (%)	100	32.32	**1.72**
Pageblocks (5473/4)	Accuracy (%)	89.77	94.19	**94.30**
	Distance	2.871	1.919	**0.652**
	Response time (ms)	**7.0**	72.6	21.2
	Activation ratio (%)	100	94.19	**2.94**
Shuttle (2175/5)	Accuracy (%)	84.51	98.62	**98.71**
	Distance	0.211	0.169	**0.009**
	Response time (ms)	**2.9**	26.1	4.6
	Activation ratio (%)	100	68.61	**4.37**
Nursery (1296/5)	Accuracy (%)	81.71	74.61	**82.10**
	Distance	18.787	**0**	**0**
	Response time (ms)	**1.8**	13.1	2.5
	Activation ratio (%)	92.27	**0.62**	8.58
Red wine (1599/6)	Accuracy (%)	57.10	58.72	**59.91**
	Distance	18.728	7.962	**5.214**
	Response time (ms)	**2.6**	21.1	10.9
	Activation ratio (%)	100	**12.46**	20.39
Yeast (1484/10)	Accuracy (%)	38.27	53.91	**54.51**
	Distance	0.003	**0**	0.001
	Response time (ms)	2.2	15.0	**1.9**
	Activation ratio (%)	100	54.45	**2.93**
Average	Accuracy (%)	73.35	80.45	**82.92**
	Distance	110.066	84.376	**2.708**
	Response time (ms)	**2.83**	22.4	5.6
	Activation ratio (%)	99.23	55.54	**9.28**

observe in average results, the 9.28% rules are activated for each input data in the proposed EBRB classification system. However, there are 99.23% activation rules in the C-EBRB. Although the 55.54% rules are activated in the DRA-EBRB, the DRA method has to determine the activation rules by recalculating activation weights for all rules many times, leading to lots of running time to classify input data.

4.3. Comparative analysis with conventional machine learning approaches

To further verify the validity of the proposed EBRB classification system, the accuracies derived from 5 independent runs with 10-fold cross-validation are further compared with conventional machine learning approaches, as shown in Table 8.

By comparing with the accuracies of these machine learning approaches, it is proved that the proposed EBRB classification system can produce satisfactory accuracies. For example, for the multi-class datasets, the 75.51% and 87.14% accuracies obtained from datasets Glass and Ecoli outperform all listed studies, especially for dataset Glass because the comparison results are based on all the machine learning approaches. For the remaining four classification datasets, although the proposed EBRB classification system fails to achieve the most satisfactory accuracies, it still improves the accuracies of the conventional EBRB system and reach the second, third, third, and second best accuracies in datasets Wine, Iris, Cancer, and Diabetes, respectively, compared to other selected approaches.

In addition, some implicit information can be further found from the comparison results:

(1) The proposed EBRB classification system has superior accuracy to handle multi-class datasets, e.g., it is clear from Table 8 that the rank of the accuracies obtained from the datasets Glass and Ecoli are better than those from the other datasets.
(2) Based on Table 8, it is almost impossible to find a best approach or system which can has the best accuracies for all classification datasets, because many factors such as data structure and noise data are likely to affect the accuracy of classifiers.

5. Conclusions and future research

In this study, a novel EBRB classification system with the new procedures of activation rule determination and weight calculation has been proposed to handle classification problems. Nineteen classification datasets were tested with 10/2-fold cross- validations to validate the efficiency and effectiveness of the proposed EBRB classification system compared with the conventional EBRB classi-

Table 8
Comparisons of accuracy for conventional machine learning approaches (the best results are highlighted in boldface).

Methods	Core supporting theory	Multi-class		Three-class		Two-class	
		Glass	Ecoli	Wine	Iris	Cancer	Diabetes
KNN [32]	K nearest neighbor	66.85%	81.27%	96.05%	85.17%	–	–
AISWNB [33]	Naive Bayes	57.74%	–	–	94.87%	**97.24%**	75.86%
EFRBCS [9]	Fuzzy set	68.57%	82.39%	–	92.00%	92.00%	71.54%
WLTSVM [5]	Support vector machine	49.91%	–	96.40%	98.00%	–	**77.08%**
LST-KSVC [34]	Neural network	65.76%	–	94.27%	**99.27%**	–	–
CMQFS [35]	Feature vector graph	70.06%	–	**98.88%**	–	–	–
HHCART [36]	Decision tree	61.90%	–	91.40%	–	97.00%	–
EBRB system [26]	Extended belief rule	51.43%	33.72%	96.32%	95.20%	94.59%	73.39%
This study	Extended belief rule	**75.51%**	**87.14%**	97.87%	95.73%	95.47%	75.98%

fication and some commonly used classification tools. The detailed contributions can be summarized into three aspects below:

(1) Different drawbacks of the conventional EBRB system, including counterintuitive individual matching degrees, insensitivity to the calculation of individual matching degrees, and the inconsistency problem have been carried out. Furthermore, a case analysis and mathematical derivation have been further applied to investigate the causes of these drawbacks, respectively.
(2) Based on the causes of the drawbacks found in the conventional EBRB system, the activation region of extended belief rules has been defined and an effective method to construct the activation region for each extended belief rule has been introduced. Besides, the new procedures of activation rule determination and weight calculation have been proposed to determine consistent activation rules and revise the calculation formula of activation weights.
(3) In the case studies of common classification datasets, the comparison results demonstrated that the proposed EBRB classification system could improve the accuracy and efficiency of the conventional EBRB systems. More importantly, for the multi-class classification datasets, the proposed EBRB classification system had a significant performance better than other improved EBRB systems and conventional machine learning approaches.

For the future research, the determination of EBRB parameters, which are often given by expert knowledge with lots of subjectivities, is one of challenges to promote the application of the EBRB system in classification problems

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Nos. 61773123, 71371053, 71501047 and 71601180), the Humanities and Social Science Foundation of the Ministry of Education under Grant (No. 14YJC630056), and the Natural Science Foundation of Fujian Province, China (No. 2015J01248).

References

[1] Z.W. Lu, L.W. Wang, Learning descriptive visual representation for image classification and annotation, Pattern Recognit. 48 (2) (2015) 498–508.
[2] P. Jaganathan, R. Kuppuchamy, A threshold fuzzy entropy based feature selection for medical database classification, Comput. Biol. Med. 43 (12) (2013) 2222–2229.
[3] P. Melin, O. Castillo, A review on type-2 fuzzy logic applications in clustering, classification and pattern recognition, Appl. Soft Comput. 21 (2014) 568–577.
[4] C.H. Tsang, S. Kwong, H. Wang, Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection, Pattern Recognit. 40 (9) (2007) 2373–2391.
[5] Y.H. Shao, W.J. Chen, Zhen, C.N. Wang, N.Y. Li, Deng, Weighted linear loss twin support vector machine for large-scale classification, Knowl. Based Syst. 73 (2015) 276–288.
[6] A. Bechini, F. Marcelloni, A. Segatori, A MapRedure solution for associative classification of big data, Inf. Sci. 332 (2016) 33–55.
[7] R. Sun, Robust reasoning: integrating rule-based and similarity-based reasoning, Artif. Intell. 75 (2) (1995) 241–295.
[8] M. Elkano, M. Galar, J. Sanz, H. Bustince, Fuzzy Rule-Based Classification Systems for multi-class problems using binary decomposition strategies: On the influence of n-dimensional overlap functions in the Fuzzy Reasoning Method, Inf. Sci. 332 (2016) 94–114.
[9] L.M. Jiao, Q. Pan, T. Denoeux, Y. Liang, X.X. Feng, Belief rule-based classification system: extension of FRBCS in belief functions framework, Inf. Sci. 309 (2015) 26–49.
[10] J.B. Yang, J. Liu, J. Wang, H.S. Sii, H.W. Wang, Belief rule-base inference methodology using the evidential reasoning approach-RIMER, IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 37 (4) (2006) 569–585.
[11] G.Y. Hu, Z.J. Zhou, B.C. Zhang, X.J. Yin, Z. Gao, Z.G. Zhou, A method for predicting the network security situation based on hidden BRB model and revised CMA-ES algorithm, Appl. Soft Comput. 48 (2016) 404–418.
[12] Y.M. Wang, L.H. Yang, Y.G. Fu, L.L. Chang, K.S. Chin, Dynamic rule adjustment approach for optimizing belief rule-base expert system, Knowl. Based Syst. 96 (2016) 40–60.
[13] J.A. Sanz, A. Fernández, H. Bustince, F. Herrera, Improving the performance of fuzzy rule-based classification system with interval-valued fuzzy sets and genetic amplitude tuning, Inf. Sci. 180 (19) (2010) 3674–3685.
[14] M.M.A. Rahhal, Y. Bazi, H. AlHichri, N. Alajlan, F. Melgani, R.R. Yager, Deep learning approach for active classification of electrocardiogram signals, Inf. Sci. 345 (2016) 340–354.
[15] A.S. Qureshi, A. Khan, A. Zammer, A. Usman, Wind power prediction using deep neural network based meta regression and transfer learning, Appl. Soft Comput. 58 (2017) 742–755.
[16] L.L. Chang, Y. Zhou, J. Jiang, M.J. Li, X.H. Zhang, Structure learning for belief rule base expert system: a comparative study, Knowl. Based Syst. 39 (2013) 159–172.
[17] Z.J. Zhou, C.H. Hu, J.B. Yang, D.L. Xu, D.H. Zhou, Online updating belief-rule-Base using the RIMER approach, IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 41 (6) (2011) 1225–1243.
[18] J.B. Yang, J. Liu, D.L. Xu, J. Wang, H.W. Wang, Optimization models for training belief-rule-based systems, IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 37 (4) (2007) 569–585.
[19] B.C. Zhang, X.X. Han, Z.J. Zhou, L. Zhang, X.J. Yin, Y.W. Chen, Construction of a new BRB based model for time series forecasting, Appl. Soft Comput. 13 (12) (2013) 3548–4556.
[20] J. Liu, L. Martinez, A. Calzada, H. Wang, A novel belief rule base representation, generation and its inference methodology, Knowl. Based Syst. 53 (2013) 129–141.
[21] L.H. Yang, Y.M. Wang, Y.X. Lan, L. Chen, Y.G. Fu, A data envelopment analysis (DEA)-based method for rule reduction in extended belief-rule-based systems, Knowl. Based Syst. 123 (2017) 174–187.
[22] L.H. Yang, Y.M. Wang, Q. Su, Y.G. Fu, K.S. Chin, Multi-attribute search framework for optimizing extended belief rule-based systems, Inf. Sci. 370-371 (2016) 159–183.
[23] J.B. Yang, D.L. Xu, On the evidential reasoning algorithm for multiple attribute decision analysis under uncertainty, IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 32 (3) (2002) 289–304.
[24] Y.M. Wang, J.B. Yang, D.L. Xu, Environmental impact assessment using the evidential reasoning approach, Eur. J. Oper. Res. 174 (2006) 1885–1913.
[25] UCI Repository of Machine Learning Database, 2018 https://archive.ics.uci.edu/ml/datasets.html.
[26] A. Calzada, J. Liu, H. Wang, K. Anil, A new dynamic rule activation method for extended belief rule-based systems, IEEE Trans. Knowl. Data Eng. 27 (4) (2015) 880–894.
[27] M. Espinilla, J. Medina, A. Calzada, J. Liu, L. Martinez, C. Nugent, Optimizing the configuration of an heterogeneous architecture of sensors for activity recognition, using the extended belief rule-based inference methodology, Microprocess Microsystems 52 (2017) 381–390.
[28] L.H. Yang, Y.M. Wang, Y.G. Fu, A consistency analysis-based rule activation method for extended belief-rule-based systems, Inf. Sci. 445-446 (2018) 50–65.

[29] Y.W. Chen, J.B. Yang, D.L. Xu, S.L. Yang, On the inference and approximation properties of belief rule based systems, Inf. Sci. 234 (2013) 121–135.
[30] J.B. Yang, Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties, Eur. J. Oper. Res. 131 (1) (2001) 31–61.
[31] S.J. Raudys, A.K. Jain, Small sample size effects in statistical pattern recognition: recommendations for practitioners, IEEE Trans. Pattern Anal. Mach. Intell. 13 (3) (1991) 252–264.
[32] J. Derrac, F. Chiclana, S. Garcia, F. Herrera, Evolutionary fuzzy k-nearest neighbors algorithm using interval-values fuzzy sets, Inf. Sci. 329 (2016) 144–163.
[33] J. Wu, S.R. Pan, X.Q. Zhu, Z.H. Cai, P. Zhang, C.Q. Zhang, Self-adaptive attribute weighting for Naive Bayes classification, Expert Syst. Appl. 42 (3) (2015) 1487–1502.
[34] Q.F. Nie, L.Z. Jin, S.M. Fei, J.Y. Ma, Neural network for multi-class classification by boosting composite stumps, Neurocomputing 149 (2015) 949–956.
[35] G.D. Zhao, Y. Wu, F.Q. Chen, J.M. Zhang, J. Bai, Effective feature selection using feature vector graph for classification, Neurocomputing 151 (2015) 376–389.
[36] D.C. Wichramarachchi, B.L. Robertson, M. Reale, C.J. Price, J. Brown, HHCART: an oblique decision tree, Comput. Stat. Data Anal. 96 (2016) 12–23.